Exploring Spatio-Temporal Variability by Eigen-Decomposition Techniques

Analisi della variabilità spazio-temporale attraverso tecniche di decomposizione spettrale

Lara Fontanella – Luigi Ippoliti
DQMTE, University `G. d'Annunzio', Viale Pindaro 42, 65127 Pescara, Italy

Kanti V. Mardia
Department of Statistics, University of Leeds, LS2 9JT, Leeds, UK

Riassunto: Lo studio della variabilità di fenomeni ambientali può essere realizzato utilizzando diverse metodologie. In questo lavoro si propone una visione unificata di una serie di tecniche multivariate che risultano particolarmente utili per l'identificazione del segnale oggetto di interesse. Facendo riferimento a processi spazialmente continui, le tecniche proposte vengono presentate nel contesto della Decomposizione Generalizzata agli Autovalori. Nonostante la metodologia considerata risulta utile per fini esplorativi, il lavoro ne propone un utilizzo anche a fini predittivi.

Key Words: Multivariate analysis, Generalized Eigenvalue Decomposition, Geostatistics, Spatio-temporal models

1. Introduction

The modelling of spatio-temporal data resulting from dynamic processes evolving in both space and time, is critical in many scientific fields such as environmental sciences, climate prediction and meteorology. Several techniques have been developed that take account of correlation in both time and space. One possible approach relies on geostatistics which provides a probabilistic framework for data analysis and predictions. For example, by including time as an additional space dimension, geostatistics extends established spatial techniques developed independently in geology (Matheron, 1962), (Journel et al., 1978), forestry (Matérn, 1980) and meteorology (Gandin, 1963), to modelling the joint spatial and temporal dependence between observations. In such cases, we only need to specify a valid spatio-temporal covariance function, and a frequently used choice is the *separable* form (Mardia and Goodall, 1993). However, notwithstanding separable covariance structure is convenient for computation and offers attractive interpretations, its form limits the nature of space-time interactions. To overcome this problem, working in a spectral domain, Cressie and Huang (1999) and Stein (2005) introduce a flexible class of nonseparable stationary covariance functions that allow for space-time interactions. Unfortunately, since the Fourier inversion can be obtained in closed form only for very special cases, the covariance function cannot always be computed explicitly.
Alternative approaches are related to models initially developed for spatial or temporal distributions. In these cases, the joint space-time dependence is often not fully modelled

nor exploited in estimation or forecasting. In particular, depending on which domain (space or time) is more densely informed, the spatio-temporal process is considered as a set of time series correlated in space (Bennet, 1979; Rhouani and Wackernagel, 1990) or as a collection of temporally correlated spatial random fields (Bogaert and Christakos, 1997). Models of the first type are relevant to cases with sparse sampling in time and dense in space, while models of the second type are suitable for cases in which data are dense in time and sparse in space. In both cases, unless some additional modelling is considered, these procedures do not allow directly for spatial and temporal predictions at locations and temporal instants for which data are not available.

In this paper, we consider multivariate data analysis techniques which are very powerful and useful for data mining. In the past, they were developed as a tool with which to explore and summarize the spatio-temporal data set, rather than as a predictive methodology (Preisendorfer and Mobley, 1988; Bretherton et al., 1992; Von Storch and Zwiers, 1999).

The motivation for exploratory methods of data analysis in the atmospheric science comes from the need of splitting the full phase space of the data into the two subspaces of *signal* and *noise*. In most general terms the signal can be a pattern in space, or in time, or in space and time, which is determined by the system dynamics. Noise, on the other hand, can be physical or instrumental and comprises all those features and details that are considered irrelevant for the signal. The identification of these two subspaces is critical for several reasons as, for example, it enables: 1) to recognizing the patterns of natural variability and distinguish them from external effects, 2) to use the physical mechanism inferred from the detected signals to construct numerical models, 3) to validate numerical models by comparing the fundamental characteristic of the modelled data with those of the observed data and finally, 4) to use the signals themselves to forecast the behavior of the system in the future.

With respect to these goals, this paper attempts first to provide a unifying overview of some statistical methods for signal detection. In particular, the signal detection is performed in the spatial context and the identification of spatial patterns of oscillations is discussed using a collection of multivariate techniques including *principal component analysis* (PCA), *canonical correlations analysis* (CCA), *partial least square* (PLS) and *redundancy analysis* (RA). Based on the Generalized Eigenvalue Decomposition (GED) (Golub and Van Loan, 1993), we provide a brief description of these techniques with the objective of highlighting similarities and differences.

The prediction problem is also of particular interest. To this purpose, we also formulate a simple parsimonious model that can be used to describe the temporal evolution in a latent space. Using a geostatistical approach, it can be shown that this model is flexible enough to allow also for spatial and spatio-temporal predictions.

The remainder of the paper is organized as follows. In Section 2 we describe the GED for spatially continuous processes, while the methodology for the detection of spatial patterns is outlined in Section 3. Section 4 describes how to obtain temporal, spatial and spatio-temporal predictions while Section 5, concludes the paper with a discussion.

2. Estimating patterns using simultaneous diagonalization

Consider a spatio-temporal process $X(\mathbf{s};t)$, where $\mathbf{s}\in D$, with D some spatial domain in two dimensional Euclidean space $\Re^2$ and $t\in\{1,2,...,T\}$ a discrete index of times. For a

given time $t=t_0$, assume that $X(\mathbf{s};t_0)$ is a zero mean second order spatial stochastic process with covariance function $Q(\mathbf{s},\mathbf{s}*)$ Then, $X(\mathbf{s};t_0)$ can be expanded in a set of deterministic functions w_k, $k\in\aleph$, which form a complete orthonormal system in the space $L_2(D)$ of square integrable functions on the domain D,

$$X(\mathbf{s};t_0)=\sum_{k\in\aleph}a_k(t_0)w_k(\mathbf{s}) \tag{1}$$

where

$$a_k(t_0)=\int_D X(\mathbf{s};t_0)w_k(\mathbf{s})d(\mathbf{s}) \tag{2}$$

are the time-dependent expansion coefficients. Both the basis functions and the expansion coefficients depend upon the analysis method used.
In general, this result is related to the probabilistic corollary of Mercer's theorem which is known as Karhunen-Loéve expansion, where the terms $w_k(\mathbf{s})$, $k\in\aleph$, are the eigenfunctions of the following homogeneous integral equation

$$\int_D Q(\mathbf{s};\mathbf{s}*)w_k(\mathbf{s})d(\mathbf{s})=\lambda_k w_k(\mathbf{s}) \qquad k\in\aleph \tag{3}$$

In this paper however, we consider an extension of the preceding expansion and discuss the case in which two kernels, $Q_1(\mathbf{s},\mathbf{s}*)$ and $Q_2(\mathbf{s},\mathbf{s}*)$, are defined. In particular, let us indicate with $Q_1(\mathbf{s},\mathbf{s}*)$ and $Q_2(\mathbf{s},\mathbf{s}*)$ two real, symmetric and square integrable functions and with Q_1 and Q_2 the integral operators with kernels $Q_1(\mathbf{s},\mathbf{s}*)$ and $Q_2(\mathbf{s},\mathbf{s}*)$. Assume also that Q_1 and Q_2 are positive definite and nonnegative definite respectively and that $\widetilde{Q}=Q_1^{-1/2}Q_2Q_1^{-1/2}$ is densely defined, bounded, and its extension to the whole of $L_2(D)$ has eigenfunctions which span $L_2(D)$. Then, if λ_k and u_k, $k\in\aleph$, are the eigenvalues and the orthonormalized eigenfunctions of $\widetilde{Q}$, we have the following simultaneous diagonalization of the two kernels (Kadota, 1967)

$$\mathrm{Q}_1(\mathbf{s},\mathbf{s}*)=\sum_{\mathrm{k}\in\aleph}\mathrm{w}_\mathrm{k}(\mathbf{s})\mathrm{w}_\mathrm{k}(\mathbf{s}*)$$
$$\mathrm{Q}_2(\mathbf{s},\mathbf{s}*)=\sum_{\mathrm{k}\in\aleph}\lambda_\mathrm{k}\mathrm{w}_\mathrm{k}(\mathbf{s})\mathrm{w}_\mathrm{k}(\mathbf{s}*)$$

where $w_k(\mathbf{s})=Q_1^{-1/2}u_k(\mathbf{s})$. The $w_k(\mathbf{s})$ also satisfies the following integral equation

$$\int_\mathrm{D}\mathrm{Q}_2(\mathbf{s},\mathbf{s}*)\mathrm{w}_\mathrm{k}(\mathbf{s})\mathrm{d}(\mathbf{s})=\lambda_\mathrm{k}\int_\mathrm{D}\mathrm{Q}_1(\mathbf{s},\mathbf{s}*)\mathrm{w}_\mathrm{k}(\mathbf{s})\mathrm{d}(\mathbf{s}) \qquad \mathrm{k}\in\aleph \tag{4}$$

which represents an extension of the Fredholm integral (3).
However, given the observed space-time series $x(\mathbf{s}_i,t)$, $i=1,\ldots,n$, $t=1,\ldots,T$, a finite approximation for equations (1),(2), (4) is required. Accordingly, we have that at each time t, the observed spatial series $\mathbf{x}(t)$ is expanded in terms of a set of n column vectors called *patterns*

$$\mathbf{x}(t) = \sum_{k=1}^{n} \tilde{\mathbf{w}}_k \tilde{a}_k(t) = \tilde{\mathbf{W}}\tilde{\mathbf{a}}(t) \tag{5}$$

where $\tilde{\mathbf{W}}$ is the *(n×n)* matrix of the patterns $\tilde{\mathbf{w}}_k$ and $\tilde{\mathbf{a}}_k$ a *(n×1) v*ector of the sample expansion coefficients obtained as a weighted linear combination of the data

$$\tilde{a}_k(t) = \sum_{k=1}^{n} \tilde{w}_k(\mathbf{s}_i) x(\mathbf{s}_i; t) = \tilde{\mathbf{w}}'_k \mathbf{x}(t) \tag{6}$$

If only a limited number *K<n* of patterns are considered, we are not able to recover the exact values of the field, thus providing a truncated expansion plus a residual $\mathbf{e}^{(K)}(t)$. In this case, equation (5) changes as follows

$$\begin{aligned} \mathbf{x}(t) &= \sum_{k=1}^{K} \tilde{\mathbf{w}}_k \tilde{a}_k(t) + \sum_{j=K+1}^{K} \tilde{\mathbf{w}}_j \tilde{a}_j(t) \\ &= \tilde{\mathbf{W}}^{(K)} \tilde{\mathbf{a}}^{(K)}(t) + \mathbf{e}^{(K)}(t) \\ &= \mathbf{x}^{(K)}(t) + \mathbf{e}^{(K)}(t) \end{aligned} \tag{7}$$

where $\tilde{\mathbf{W}}^{(K)}$ is a *(n×K)* matrix, $\tilde{\mathbf{a}}^{(K)}(t)$ is a *(K×1)* vector and $\mathbf{x}^{(K)}(t)$ is a process smoother than $\mathbf{x}(t)$.

The simultaneous diagonalization of the two kernels can be approximated as

$$\sum_{k=1}^{n} \tilde{\mathbf{Q}}_2(\mathbf{s}_i, \mathbf{s}_j) w_k(\mathbf{s}_i) = \tilde{\lambda}_k \sum_{k=1}^{n} \tilde{\mathbf{Q}}_1(\mathbf{s}_i, \mathbf{s}_j) w_k(\mathbf{s}_i) \tag{8}$$

or in matrix formulation

$$\tilde{\mathbf{Q}}_2 \tilde{\mathbf{w}}_k = \tilde{\lambda}_k \tilde{\mathbf{Q}}_1 \tilde{\mathbf{w}}_k \tag{9}$$

where $\tilde{\mathbf{Q}}_2$ and $\tilde{\mathbf{Q}}_1$ are the *(n×n)* matrices of the values of the kernel functions $Q_1(\mathbf{s},\mathbf{s}^*)$ and $Q_2(\mathbf{s},\mathbf{s}^*)$ onto the spatial sample points. Equation (9) constitutes a *generalized eigenvalue decomposition* (GED) (Golub and Van Loan, 1993).

If $\tilde{\mathbf{Q}}_2$ and $\tilde{\mathbf{Q}}_1$ are symmetric and $\tilde{\mathbf{Q}}_1$ is positive definite, then the eigenvalues $\tilde{\lambda}_k$ and the eigenvectors $\tilde{\mathbf{w}}_k$ are real. Further, if the eigenvalues are distinct, the different eigenvectors are orthogonal in the metrics $\tilde{\mathbf{Q}}_2$ and $\tilde{\mathbf{Q}}_1$

$$\tilde{\mathbf{W}}' \tilde{\mathbf{Q}}_1 \tilde{\mathbf{W}} = \mathbf{I} \tag{10}$$

$$\tilde{\mathbf{W}}' \tilde{\mathbf{Q}}_2 \tilde{\mathbf{W}} = \tilde{\mathbf{\Lambda}} \tag{11}$$

where $\tilde{\mathbf{\Lambda}}$ is a diagonal matrix.

Finally, we also notice that if the matrix $\widetilde{\mathbf{Q}}_1$ is positive definite, equation (9) can be handled by the equivalent expression

$$\widetilde{\mathbf{Q}}_1^{-1}\widetilde{\mathbf{Q}}_2\widetilde{\mathbf{w}}_k = \widetilde{\lambda}_k\widetilde{\mathbf{w}}_k \tag{12}$$

In this case, the matrix $\widetilde{\mathbf{Q}}_1^{-1}\widetilde{\mathbf{Q}}_2$ is generally not symmetric, but it is possible to recover a symmetric eigenvalue problem using, for example, the Cholesky decomposition, $\widetilde{\mathbf{Q}}_1 = \mathbf{L}\mathbf{L}'$, and considering the eigenvalue decomposition of the symmetric matrix $\widetilde{\mathbf{Q}} = \mathbf{L}^{-1}\widetilde{\mathbf{Q}}_2(\mathbf{L}^{-1})'$. Its eigenvalues are the same of the original problem, while its eigenvectors are $\widetilde{\mathbf{u}}_k = \mathbf{L}'\widetilde{\mathbf{w}}_k$.

3. Detection of spatial patterns

The aim of the techniques of exploratory data analysis is to summarize the dominant characteristics of a field, such as the dominant space patterns, and to discriminate between the signal of interest and the unrelated processes or noise. Methods of spatial pattern detection attempt to exploit the information available in spatial distributed data and involve eigenvalue decomposition in the form of (9). In fact, as noted in Section 2, at a given time *t*, the observed spatial series **x**(t) can be (approximately) described in the basis vectors spanning a (local) subspace in terms of projections onto the new basis vectors. The criterion for the selection of the new basis vectors is dependent on the application.
In the following paragraphs, we shall present four different criteria that emerge as solutions to special cases of the generalized eigenproblem.

3.1 Principal component analysis or EOF

Principal component analysis (PCA) provides a widely used method of describing patterns of variables observed over a large spatial area. Most of applications appear in meteorological and climatological journals where the technique is more frequently known as Empirical Orthogonal Functions (EOFs). However, Richman (1986) distinguishes between EOF analysis and PCA, with the former having unit-length eigenvectors and the latter having eigenvectors renormalized to have lengths proportional to their respective eigenvalues. In this paper, the terms PCA and EOF analysis are used as synonymous.
The goal of EOF analysis is to provide a compact description of the spatial variability of data series in terms of orthogonal functions or statistical "modes". They are called empirical to reflect the fact that they are defined by the covariance structure of the observed data set.
There are two approaches for computing EOF. The first constructs the covariance matrix of the data series and then decompose it into eigenvalues and eigenvectors; the second uses the singular value decomposition of the data matrix to obtain eigenvalues, eigenvectors and time varying amplitudes (principal components). The EOF obtained from the two methods are identical if the spatial covariance structure is computed by

using the standard method of moments (MOM), $\mathbf{C}=(T-1)^{-1}\mathbf{X}'\mathbf{X}$, where the ($T\times n$) data matrix $\mathbf{X}$ is assumed to be column centered. This does not hold, for example, if a geostatistical approach is considered to obtain the empirical spatial covariance matrix.
The spectral decomposition of the symmetric covariance matrix $\mathbf{C}$ is a special case of the generalized eigenvalue decomposition (9), where $\widetilde{\mathbf{Q}}_2 = \mathbf{C}$ and $\widetilde{\mathbf{Q}}_1 = \mathbf{I}$. In this case, the k-th column of matrix $\widetilde{\mathbf{W}}$ given by $\widetilde{\mathbf{w}}_k \equiv (\widetilde{w}_k(\mathbf{s}_1),...,\widetilde{w}_k(\mathbf{s}_n))'$, represents the k-th EOF associated to the corresponding expansion coefficient, $\widetilde{a}_k(t)$, known as principal component. Sorting the EOFs according to the magnitude of their eigenvalues, the process can be reconstructed following equation (5). Notice that, if a truncation level K is chosen such that the process is reconstructed through equation (7), then PCA gives a data dependent set of basis vectors that is optimal in statistical mean square sense.
An objection of the use of PCA is that it does not exploit the order of the data, in the sense that the patterns (EOFs) used to form the principal components are invariant to permutation of order. Thus, a number of alternative approaches have been developed to incorporate temporal information.

3.2 Partial least square

The partial least square allows for the identification of pairs of spatial patterns and time coefficients which account for a fraction of the covariance between two processes analyzed jointly. Given two ($T\times n$) data matrices, $\mathbf{X}$ and $\mathbf{Y}$, of zero-mean spatiotemporal series $x(\mathbf{s}_i;t)$ and $y(\mathbf{s}_i;t)$, i=1,…,n, t=1,…,T, with between-sets spatial covariance matrix $\mathbf{C}_{xy}$, the goal is to find the two directions of maximal data covariation; *i.e* the directions $\widetilde{\mathbf{w}}_{x_k}$ and $\widetilde{\mathbf{w}}_{y_k}$ such that the expansion coefficients, $\widetilde{\mathbf{a}}_{x_k} = \mathbf{X}\widetilde{\mathbf{w}}_{x_k}$ and $\widetilde{\mathbf{a}}_{y_k} = \mathbf{Y}\widetilde{\mathbf{w}}_{y_k}$, have maximum covariance. Following de Jong (1986), the first two spatial patterns, $\widetilde{\mathbf{w}}_{x_1}$ and $\widetilde{\mathbf{w}}_{y_1}$, are identified by maximizing the covariance of the expansion coefficients, given by $\mathrm{cov}(\widetilde{\mathbf{a}}_{x_1}, \widetilde{\mathbf{a}}_{y_1}) = \widetilde{\mathbf{w}}'_{x_1}\mathbf{C}_{xy}\widetilde{\mathbf{w}}_{y_1}$, subject to the constraint $\widetilde{\mathbf{w}}'_{x_1}\widetilde{\mathbf{w}}_{x_1} = \widetilde{\mathbf{w}}'_{y_1}\widetilde{\mathbf{w}}_{y_1} = 1$. In this case, it can be shown that the procedure leads to the pair of eigenvalue problems

$$\begin{aligned} \mathbf{C}_{xy}\widetilde{\mathbf{w}}_{y_1} &= \widetilde{\lambda}_1\widetilde{\mathbf{w}}_{x_1} \\ \mathbf{C}_{yx}\widetilde{\mathbf{w}}_{x_1} &= \widetilde{\lambda}_1\widetilde{\mathbf{w}}_{y_1} \end{aligned} \tag{13}$$

that can be solved by the generalized eigenvalue decomposition (9) setting

$$\widetilde{\mathbf{Q}}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{0} \end{bmatrix} \qquad \widetilde{\mathbf{Q}}_1 = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

Then, the remaining set of expansion coefficients are the linear combinations $\widetilde{\mathbf{a}}_{x_k} = \mathbf{X}\widetilde{\mathbf{w}}_{x_k}$ and $\widetilde{\mathbf{a}}_{y_k} = \mathbf{Y}\widetilde{\mathbf{w}}_{y_k}$, that are uncorrelated with the previous k-1 pairs and maximize the covariance.
It can be shown that the computation of the patterns of maximum spatial covariance can also be obtained by finding the singular value decomposition of the cross-covariance matrix. For this reason in meteorological and climatological studies the PLS technique

is also known as Singular Value Decomposition of coupled fields and is equivalent to the Combined EOF analysis of two variables (Bretherton et al., 1992).

3.3 Canonical correlation analysis

Another multivariate statistical method that has been widely used for the identification of spatial patterns is *canonical correlation analysis* (CCA).
As in PLS two processes are considered, but the goal here is to find the two directions of maximal data correlation, i.e the direction $\widetilde{\mathbf{w}}_{x_k}$ and $\widetilde{\mathbf{w}}_{y_k}$ such that the linear combination $\widetilde{\mathbf{a}}_{x_k} = \mathbf{X}\widetilde{\mathbf{w}}_{x_k}$ and $\widetilde{\mathbf{a}}_{y_k} = \mathbf{Y}\widetilde{\mathbf{w}}_{y_k}$, have the largest possible correlation (Mardia et al., 1979). The first spatial patterns $\widetilde{\mathbf{w}}_{x_1}$ and $\widetilde{\mathbf{w}}_{y_1}$, are thus identified by maximizing the correllation of the expansion coefficients $\widetilde{\mathbf{a}}_{x_1}$ and $\widetilde{\mathbf{a}}_{y_1}$, given by

$$corr\left(\widetilde{\mathbf{a}}_{x_1}, \widetilde{\mathbf{a}}_{y_1}\right) = \frac{\widetilde{\mathbf{w}}'_{x_1} \mathbf{C}_{xy} \widetilde{\mathbf{w}}_{y_1}}{\sqrt{\left(\widetilde{\mathbf{w}}'_{x_1} \mathbf{C}_{xx} \widetilde{\mathbf{w}}_{x_1}\right)\left(\widetilde{\mathbf{w}}'_{y_1} \mathbf{C}_{yy} \widetilde{\mathbf{w}}_{y_1}\right)}} \tag{14}$$

The maximum of (14), subject to $\widetilde{\mathbf{w}}'_{x_1} \mathbf{C}_{xx} \widetilde{\mathbf{w}}_{x_1} = \widetilde{\mathbf{w}}'_{y_1} \mathbf{C}_{yy} \widetilde{\mathbf{w}}_{y_1} = 1$, leads to the pair of eigenvalue problems

$$\begin{aligned} \mathbf{C}_{xy} \widetilde{\mathbf{w}}_{y_1} &= \widetilde{\lambda}_1 \mathbf{C}_{xx} \widetilde{\mathbf{w}}_{x_1} \\ \mathbf{C}_{yx} \widetilde{\mathbf{w}}_{x_1} &= \widetilde{\lambda}_1 \mathbf{C}_{yy} \widetilde{\mathbf{w}}_{y_1} \end{aligned} \tag{15}$$

that can be solved by the generalized eigenvalue decomposition (9) setting

$$\widetilde{\mathbf{Q}}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{0} \end{bmatrix} \qquad \widetilde{\mathbf{Q}}_1 = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{yy} \end{bmatrix}$$

Of course, also in this case the remaining set of expansion coefficients are the linear $\widetilde{\mathbf{a}}_{x_k} = \mathbf{X}\widetilde{\mathbf{w}}_{x_k}$ and $\widetilde{\mathbf{a}}_{y_k} = \mathbf{Y}\widetilde{\mathbf{w}}_{y_k}$, that are uncorrelated with the previous *k*-1 canonical pairs, have unit variance, and maximize the related correlation as in (14).

3.4 Redundancy analysis

Given two processes, if the aim is to predict one process as well as possible in the least square error sense, the spatial patterns must be chosen so that this error measure is minimized. This corresponds to a low-rank approximation of multivariate linear regression, which is also known as reduced rank regression (Izenman, 1975) or as redundancy analysis (RA) (van de Wollenberg, 1977).
RA was first described in the psychometric literature, but has only recently been applied in environmental studies (Von Storch and Zwiers, 1999). It is a technique that is used to associate patterns of variation in a predictor field with patterns in the predictand field. Thus, differently from CCA and PLS, it treats the **X** and **Y** processes asymmetrically. In

particular, RA seeks to find pairs of predictor and predictand patterns that maximize the predictand variance and this is directly addressed by identifying patterns that are strongly related through the most efficient multivariate regression on **Y**.
To measure the degree to which one set of variables **X** can predict the other set of variables **Y**, the *redundancy index* can be used

$$R^2(\mathbf{Y}:\mathbf{X}) = \frac{tr(\mathbf{C}_{\hat{y}\hat{y}})}{tr(\mathbf{C}_{yy})} = \frac{tr(\mathbf{C}_{yx}\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy})}{tr(\mathbf{C}_{yy})} \tag{16}$$

where *tr* denote the trace of the matrix. This index represents the proportion of the total variance in **Y** that can be accounted for by the linear regression of **Y** on **X**. The redundancy index is invariant under orthogonal transformations of the predictand variables **Y** and under non-singular transformations of the predictor variables **X**. Furthermore it has the property of decomposability over any complete set of uncorrelated linear combinations of the predictor variables **X**; that is if $\widetilde{\mathbf{w}}_{x_1}, \widetilde{\mathbf{w}}_{x_2}, ..., \widetilde{\mathbf{w}}_{x_n}$ is any set of non-zero vectors, such that $\widetilde{\mathbf{w}}'_{x_i}\mathbf{C}_{xx}\widetilde{\mathbf{w}}_{x_j} = 0$ for $i \neq j$, then $R^2(\mathbf{Y}:\mathbf{X}) = \sum_{j=1}^{n} R^2(\mathbf{Y}:\mathbf{X}\widetilde{\mathbf{w}}_{x_j})$.
In practice, it can be shown that the maximization of the redundancy index, and hence the identification of the best predicted and predictor patterns, is related to the solution of the following pair of eigenvalue problems

$$\begin{aligned} \mathbf{C}_{xy}\widetilde{\mathbf{w}}_{y_1} &= \widetilde{\lambda}_1\mathbf{C}_{xx}\widetilde{\mathbf{w}}_{x_1} \\ \mathbf{C}_{yx}\widetilde{\mathbf{w}}_{x_1} &= \widetilde{\lambda}_1\widetilde{\mathbf{w}}_{y_1} \end{aligned} \tag{17}$$

that can be solved by the generalized eigenvalue decomposition (9) setting

$$\widetilde{\mathbf{Q}}_2 = \begin{bmatrix} \mathbf{0} & \mathbf{C}_{xy} \\ \mathbf{C}_{yx} & \mathbf{0} \end{bmatrix} \qquad \widetilde{\mathbf{Q}}_1 = \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

4. A Simple prediction model

In this section, dealing with the prediction problem, we develop a simple and parsimonious model which accounts for the main features of the data.
Assuming two processes **X** and **Y** are observed, of particular interest is the *asymmetric* case, where it is explicitly recognized that one variable can be used to predict the other variable. In such a case, letting **Y** the process to be predicted, we can calculate the least squares estimate of the first *K* time series generated by the expansion coefficients $\widetilde{\mathbf{a}}_y^{(K)}(t)$, given $\widetilde{\mathbf{a}}_x^{(K)}(t)$, as

$$\widetilde{\mathbf{a}}_y^{(K)}(t) = \mathbf{B}\widetilde{\mathbf{a}}_x^{(K)}(t) + \mathbf{v}(t) \qquad K = 1,...,n \tag{18}$$

where **B** is a ($K\times K$) matrix of regression coefficients and **v**(t) a ($K\times 1$) vector of zero mean independent error terms with variance σ_v^2. Then, we can specify **Y** by using a truncated set of basis functions (*principal fields*) as

$$\hat{\mathbf{y}}^{(K)}(t) = \sum_{k=1}^{K} \widetilde{\mathbf{w}}_{y_k} \hat{\widetilde{a}}_{y_k}(t) = \widetilde{\mathbf{W}}_y^{(K)} \hat{\widetilde{\mathbf{a}}}_y^{(K)}(t) \qquad K = 1,...,n \tag{19}$$

where $\hat{\widetilde{\mathbf{a}}}_y^{(K)}(t) = \mathbf{B}\widetilde{\mathbf{a}}_x^{(K)}(t)$.

It can be shown that equations (18-19) represent a parsimonious model that can be easily used to get temporal, spatial and spatio-temporal predictions. The number of spatial oscillation patterns, *K*, acts as a *regularization* parameter. For *K* sufficiently small, only a few number of expansion coefficients have to be considered in the analysis.

For $t \le T$ the model can be used to fit the **Y** data or, eventually, to reconstruct missing data. On the other hand, further details must be provided if we are trying to predict the process at unobserved spatial locations. In particular, if such a prediction is required at time *t*, $t \le T$, and at an unmonitored site $\mathbf{s}_0$, a straightforward approach might use the following equation

$$\hat{y}(\mathbf{s}_0; t) = \sum_{k=1}^{K} \hat{w}_{y_k}(\mathbf{s}_0) \hat{\widetilde{a}}_{y_k}(t) \tag{20}$$

where $\hat{w}_{y_k}(\mathbf{s}_0)$ is the predicted *principal fields* at site $\mathbf{s}_0$. The prediction of $\hat{w}_{y_k}(\cdot)$ is not a difficult task and, ensuring orthogonality, we could apply some relatively simple interpolation schemes. To this end, Mardia and collegues (1998) and Wikle and Cressie (1999), provide two alternative approaches.

For $t>T$, the model described by equations (18-19) also allows for temporal predictions. In fact, assuming that the independent variable is available (observable) for the forecasting period or that, using for example ARMA or VARMA models (Hamilton, 1994), it is possible to obtain "good" forecasts of *X*, it is straightforward to get temporal or spatio-temporal predictions for *Y*.

Considering now the role played by specific techniques in the context of prediction, it might be worth to focus on the estimation of the regression matrix **B**. It is well known that, in general

$$\mathbf{B} = E\left[\widetilde{\mathbf{a}}_x(t)\widetilde{\mathbf{a}}_x(t)'\right]^{-1} E\left[\widetilde{\mathbf{a}}_x(t)\widetilde{\mathbf{a}}_y(t)'\right].$$

A point worth noting is that, because of the constraint $\widetilde{\mathbf{W}}_x'\mathbf{C}_{xx}\widetilde{\mathbf{W}}_x = \mathbf{I}$, both for RA and CCA, the estimated regression matrix is $\hat{\mathbf{B}} = \widetilde{\mathbf{W}}_y'\mathbf{C}_{yx}\widetilde{\mathbf{W}}_x$. In particular, it is straightforward to show that it is diagonal and that the parameters are equal to the eigenvalues $\widetilde{\lambda}_k$, k=1,…,*K*, estimated by the generalized decomposition. Then, considering also that from the constraint we have $\widetilde{\mathbf{W}}_x = \left(\widetilde{\mathbf{W}}_x'\mathbf{C}_{xx}\right)^{-1}$, from (19) it follows

$$\hat{\mathbf{y}}(t) = \tilde{\mathbf{W}}_y \left(\tilde{\mathbf{W}}_y' \mathbf{C}_{yx} \tilde{\mathbf{W}}_x\right) \tilde{\mathbf{a}}_x(t) = \tilde{\mathbf{W}}_y \left(\tilde{\mathbf{W}}_y' \mathbf{C}_{yx} \tilde{\mathbf{W}}_x\right) \tilde{\mathbf{W}}_x' \mathbf{x}(t) = \mathbf{C}_{yx} \mathbf{C}_{xx}^{-1} \mathbf{x}(t) \qquad (21)$$

from which, it can be noted that, if all the expansion coefficients are used (*i.e.* K=n) and one single field is considered, such that $\mathbf{y}(t)=\mathbf{x}(t+\tau)$ with τ a temporal lag, predictions based on CCA or RA are *identical* to those based on vectorial autoregressive models (VAR) (Hamilton, 1994).

5. Conclusion

In this paper, we have proposed a unifying overview of a set of multivariate data analysis techniques which are very powerful and useful for signal detection. In the context of spatially continuous processes, all the techniques have been presented in terms of Generalized Eigenvalue Decomposition (GED). A point worth noting here is that as in the case of the Karhunen-Loéve expansion, the difficulties of the approach are considerable for a continuous domain when data are collected only from a sparse and irregular network. The fact that we are considering a process observed at discrete points is a practical limitation to the numerical solution of the mentioned equations. Accordingly, if there are n sample points in the domain, only n eigenfunctions can be estimated while, indeed, there are a denumerable infinity for a continuous process. Thus, the geometrical relations involving the domain of integration and the relations between the sites $\mathbf{s}_i$, i=1,…,n are completely ignored in a "discrete" matrix formulation. However, this limitation should be recognized as a restriction on the accuracy of the solution, but not as a part of the problem formulation. Hence, the numerical problem encountered in practice, is to estimate the covariance kernels and attempt to solve equation (1), (2) and (4). One simple solution to this problem is represented by a gridding procedure. In this case, by using some simple predictor, the field is predicted at regular grid locations so that the basis functions are calculated from this field defined on the regular grid (Karl et al., 1982). Alternatively, following the same lines of the approaches proposed by Cohen and Jones (1969), Buell (1972) and Obled and Creutin (1986), numerical quadrature solutions could also be considered.

The similarities and differences between the four methods can be seen by comparing the matrices $\tilde{\mathbf{Q}}_1$ and $\tilde{\mathbf{Q}}_2$ in the generalized eigenproblem. PCA preserves as much variance as possible given a certain dimensionality of the model. It also only concerns one set of variables while the other three, in principle, concern relations between two sets of variables. In applications where relations between two sets of data are considered, the basis vectors can be derived by considering the maximization either of the data covariation, as in the PLS, or the data correlation as in the CCA. Furthermore, while in CCA, the between-sets covariance matrices are normalized with respect to the within-set covariances in both the X and the Y spaces, in PLS no normalization is done. If the goal is to predict a signal as well as possible in the least square error sense, the basis might be chosen by using the RA technique. RA differs from the other three problems in that it is formulated as a mean square error problem, while the other three methods are formulated as maximization problems. Further, in RA the normalization is done only with respect to the X space covariance while the Y space, where the square error is defined, is left unchanged. In any case, these three cases can be seen as the same

problem, covariance maximization, where the variables have been subjected to different, data dependent, scaling.
Also when only one field is analyzed, the three techniques (PLS, CCA and RA) can still be used. In particular, if one is interested in studying the temporal dynamic of the field, they might be considered to explore the relationship between $\mathbf{x}(t)$ and $\mathbf{y}(t)=\mathbf{x}(t+\tau)$. In this case, for example, if CCA is used to find the canonical correlation patterns between $\mathbf{x}(t)$ and $\mathbf{x}(t+\tau)$, the columns of $\widetilde{\mathbf{W}}_x$ will represent spatial patterns at time t that are correlated with spatial patterns $\widetilde{\mathbf{W}}_y$ at time $t+\tau$. This is useful above all for temporal prediction applications as the presence of a pattern $\widetilde{\mathbf{W}}_x$ at given time t, indicates that it is likely that the pattern $\widetilde{\mathbf{W}}_y$ will emerge τ time units later.

References

Bennett R.J. (1979) *Spatial Time Series: Analysis -Forecasting – Control*, London, Pion Limited

Bogaert P., G. Christakos, (1997) Stochastic Analysis of Spatiotemporal Solute Content Measurements Using a Regressive Model. *Stochastic Hydrology and Hydraulics,* 11, 267-295.

Bretherton C. S., Smith C., Wallace J. M. (1992) An intercomparison of methods for finding coupled patterns in climate data. *J. Climate,* 5, 541-560.

Buell C.E. (1972) Integral Equation Representation for Factor Analysis. *J. Atmosph. Sci.*, 28, 1502-1505

Cohen A., Jones R.H. (1969) Regression on a Random Field, *JASA*, Vol. 64, pp. 1172-1182

Cressie N., Huang H.L. (1999) Classes of non-separable, spatio-temporal stationary covariance functions, *JASA*, 94, 1330-1340

de Jong S. (1983) Simpls: an alternative approach to partial least squares regression, *Chemom. Intell. Lab. Systems*}, 18, 251-263

Gandin L. (1963) *Objective Analysis of Meteorological Fields: Gidrometeorologicheskoe Izdatel'stvo (GIMEZ),* Leningrad. Reprinted by Israel Program for Scientific Translations, Jerusalem, 1965

Golub G.H., Van Loan C.F. (1993) *Matrix Computations*, John Hopkins University Press

Hamilton J.D. (1994) *Time Series Analysis*, Princeton University Press, Princeton

Izenman A.J. (1975) Reduced-rank regression for the multivariate linear model, *Journal of Multivariate Analysis*, 5 248-264

Journel A.G., Huijbregts Ch. J. (1978) *Mining Geostatistics,* Academic Press, New York

Kadota T.T. (1967) Simultaneous Diagonalization of two Covariance Kernels and Application to Second Order Stochastic Processes, *SIAM J. Appl. Math.*, 15, N. 6, 1470-1480

Karl T.R., Koscielny A.J., Diaz H.F. (1982) Potential errors in the application of principal components (eigenvector) analysis to geophisical data. *Appl. Meteor.*, 21, 1183-1186

Mardia K.V., Goodall C.R. (1993) Spatial-temporal Analysis of multivariate environmental data, in *Patil G.P. and Rao C.R. Eds. Multivariate Environmental Statistics*; Elsevier, Amsterdam, p.347-386

Mardia K.V., Redfern E., Goodall C.R., Alonso F. (1998) The Kriged Kalman Filter. *TEST*, 7, N. 2, 217-285

Mardia K.V., Kent J.T., Bibbi J.M. (1979) *Multivariate Analysis*, Academic press, London

Matérn B. (1980) *Spatial variation,* Lecture Notes in Statistics, 36, Springer Verlag, New York, 2nd Ed.

Matheron G. (1962) *Traité de géostatistique appliqué,* Tome I: Mémoires du Bureau de Recherches Géologiques et Miniéres, Editions Technip, Paris

Obled C., Creutin J.D. (1986) Some Developments in the Use of Empirical Orthogonal Functions for mapping Meteorological Fields, Journal of Climate and Applied meteorology, 25, N. 9, 1189-1204

Preisendorfer R.W., Mobley C.D. (1988) *Principal Component Analysis in Meteorology and Oceanography*, Amsterdam, Elsevier

Richman M.B. (1986) Rotation of Principal Components, *J. Climatol*., 6, 293-335

Rhouani S., Wackernagel H. (1990) Multivariate geostatistical approach to space-time data analysis, *Water Resources*, 26, N. 4, 585-591

Stein, M. L. (2005), Space-time covariance functions, *JASA*, 100, 310-321.

Von Storch H., Zwiers F.W. (1999) *Statistical Analysis in Climate Research*, Cambridge University Press

Wikle C. K., Cressie, N. (1999) A Dimension-Reduction Approach to Space-Time Kalman Filtering, *Biometrika*, 86, N.4, 815-829

van de Wollenberg A.L. (1977) Redundancy analysis:an alternative for canonical correlation analysis, *Psycometrica*, 36, 207-209